比特币交易可视分析

Category: Research

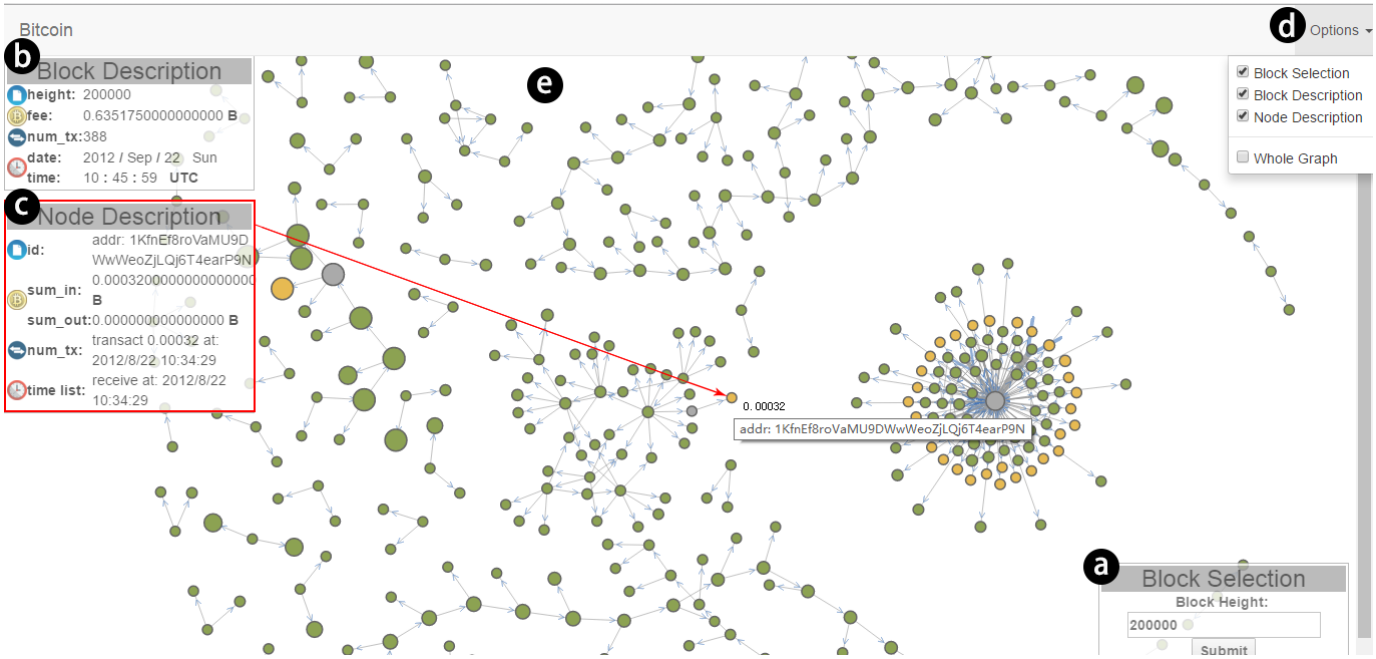

Fig. 1. 针对比特币交易可视分析系统界面，(a)选择输入框，(b)块信息，(c)节点信息，(d)选择显示提示栏，(e)交易网络

Abstract—比特币近年来已经在成为最重要的加密货币。相对于其他电子货币来说，它最重要的就是匿名性。比特币的一切的交易的详细记录都保存在一个公共账本上。用地址表示交易的发送方和接收方，通过公钥来确认交易。虽然交易记录公开，但是我们无法确认交易背后的人。比特币的匿名性在商业领域引起广泛的用途，但是也引起了人们的担忧。比特币的交易网络可以看成是一个拥有超过7千万节点的有向图，每个点用来表示一个交易，每个边用来表示交易之间的关系。我们构建了一个可视化系统用来查看比特币的交易。通过分析其交易探索其匿名性和交易之间关联。通过这个系统我们可以发现总结出用户的交易模式，发现一些异常的交易行为。

Index Terms—比特币, 电子货币, 图可视化, 可视分析

1 简介

近年来比特币作为电子货币用途越来越广。我们想要对传统的金融数据进行分析，但是很难找到公开的数据。对于电子货币中的最重要的货币比特币来说，它的历史交易数据是完全公开的，给我们提供了一个机会去研究和分析交易数据。整个交易数据组成了一个网状的图结构，所以我们可以采用图分析的方法来帮组我们探索数据。比特币具有很多金融数据所没有的特性，具有很强的匿名性，没有一个中央的管理机构，其交易的认证依赖于整个P2P网络，通过这种机制可以保证交易数据的安全，而且能够极大的降低其交易的维护费用，目前已经有部分国家承认或者即将承认比特币的合法地位，比特币在金融搜索热词中排名第二仅次于支付宝。由比特币产生的区块链技术，目前正准备应用在银行交易的管理上。央行更提出了要发行数字货币。足以看出，比特币的重要程度，电子货币是未来的发展方向。但是过于隐私的交易也会带来不安全的因素，丝路（Silk Road）从发展繁荣到最后被查封，足已说明了，比特币巨大的影响力，同时也会带来隐患。不法分子会利用比特币的进行黑色交易。在这个大背景下，对比特币的交易的分析显得尤为重要。

2 比特币的机制

比特币是一个共识网络，它是有史以来的首个没有中心的对等支付网络货币。它产生了一个崭新的付款系统和一种真正数字化的货币。用户完全可以自己掌控而不需要额外的中间人或中央管理机构。从大众的角度来看，比特币就像互联网货币。比特币和传统货币有很大的区别，比特币并不依赖一个可信的实体机构，例如银行政府等。比特币的发布和交易是依靠整个P2P网络 [17]，没人任何人控制整个比特币网络，当所有的用户达成共识时，整个比特币网络才可以正常的运作。比特币的所有交易存在同一个账本中，该账本称为块链（block chain）。块链是由块（block）所组成的链表构成的。比特币的交易是地址（address）来进行的。

3 相关研究工作

3.1 图可视化

图广泛用于描述物体之间的差别和关系。典型的使用在表示社会网络和交通网络上 [11]。虽然可以使用图来表示出所有的元素，很少很难分清和分析数据，如果图过于庞大。NicheWorks [20], GVF [14]和H3Viewer [15]都是用来处理大数据集。在处理大量数据的时候，没有一种算法能够保证布局的合理性，我们甚至没有意义将它们都表示出来 [10]。我们可以将传统的布局算法结合起来去使用。

Eades [6] [1]首次提出在图布局算法中，使用弹簧模型。使用胡克定律(Hooke's law)来描述实体之间的力。在这个算法中，他引入了引力和斥力的概念。如果两个顶点之间存在边，则具有引力。对于任何两个顶点都存在斥力。弹簧模型在提出后被广泛的使用，之后也出现了许多弹簧模型的变种。例如KK模

型 [12]提出了两个顶点之间的理想间距和它们中的最短的路径的距离成正相关。还有FR模型 [9]结合了天体重力模型与胡克定律(Hook'S law)。弹簧模型可以产生一个平衡稳定的布局，不用添加额外的条件 [8]，它能保证边的交叉最小。但是这种方法时间复杂度高达O (N3)，其中N为节点的数目。而且稳定性差，两次运行算法可能会产生不同的布局。

绝大部分的图的布局仅仅由边和点决定的。但是有时候我们需要在平面预先指定点的位置。例如对于附带地理信息坐标的数据，我们需要通过使用折线或者样条曲线来画边 [3] [4] [16]。

3.2 比特币可视分析

绝大多数的研究集中在对比特币网络的分析，和追踪大额交易的流向，从而发现一些异常的交易和总结出一些交易模式，帮组我们探索交易的过程，查看交易之间的联系。

Battist等人 [2]构建了一个可视化分析系统，用交易图表示比特币流如何和何时与其他流的混合过程。他们提出了比特币的纯净度，可以让用户理解比特币何时和怎样混入交易的。他们分析了单个和多个交易的比特币的流向，能够让我们的看到，每一笔金额的详细流通过程。但是他们的表示方法不是那么直观。

Reid和Harrigan [18]将属于同一个发送方的所有地址合并为一个实体，然后分析合并后的结果。Fleder分析了比特币交易图，主要为了揭开比特币的隐私性，他们提出了很多关于比特币机制后面所隐藏的经济学规则 [7]。

Ron和Shamir [19]对比特币交易的历史数据进行了大量的统计得出了很多有意义的信息。他们发现有些机构包含大量的比特币地址，并且统计了他们所拥有的比特币的数目。绝大部分%以上的地址所拥有的比特币数目小于1比特币，这个还是可以理解，这个与比特币官方推荐每次交易使用新的交易地址有关。他们还总结除了很多的交易模式。有一个长链的交易，在每个链接中，每个地址在传递比特币的时候，都会留下相同数量的比特币，然后将剩下的比特币继续向后面传递。存在很多地址，直接或者间接向同一个目的地址发送比特币。还有一个实体首先将90,000 BTC发送到其他地址，那些地址又将这些比特币发回这个实体，所有的这些操作在一天之内完成。我们可以看到这些交易存在奇怪的关系，这个交易在尝试着隐藏某些信息，所以他们极有可能是在洗钱。还有些交易分割成二叉树的模式。这些异常的交易是我们所研究的重点，我们希望分析这些交易给出合理的解释，分析交易背后人的行为模式，目的是为了揭露比特币的匿名性，探索虚拟货币与现实世界的联系。

很多人喜欢在比特币最大的论坛bitcointalk上发帖并且附带上自己的比特币地址要求捐献。虽然比特币官方声明不要泄露自己的地址，因为这样别人可以知道谁拥有了这些地址，会破坏比特币的匿名性。但是很多人并不在意。Fleder等人 [7]的工作就是通过爬虫这个论坛获取到大量的地址，然后很容易可以获取到地址与用户名的对应关系。然后他们分析经常在这个论坛上活动的人的交易的关系。从下图中我们可以看出存在一些大额的交易和社区。从而我们可以分析，对于那些经常交易的人，他们之间应该存在某种关系。

Lopez等人 [13]从bitcointalk论坛上爬取不同的主题类型的帖子，然后使用SVM（支持向量机）对帖子进行分类，分析情感倾向，然后根据这个预测比特币与真实货币的汇率。

3.3 比特币的地址合并

比特币的交易比较特殊。通常有多个发送方和多个接收方。当一个人给另一个转账时，他的每个地址中所拥有的比特币都小于支付金额时，可以将多个地址中的比特币放在一块进行发送，这就出现了发送地址为多个的情况。一个人可以同时给多个人转账，而且通过地址转账，会将找零存入发送方的地址中，也就是说，接收方中的地址中可能存在发送方的地址。还有些证据表明，有些机构为了增强交易的混淆性和不可追踪性，会将多个交易合并成一个交易，但是这种情况很少。所以一般可以认为所有发送方地址应该属于同一个人。如果一笔交易的发送方地址包括地址a和b，另一笔交易的发送方地址包括地址b和c，那么我们可以认为地址a，b和c属于同一个实体。我们还可以通过并查集（the Union-Find） [5]来计算交易地址的传递闭包或者一些聚类算法来合并地址。

4 系统介绍

本文主要介绍比特币交易地址合并的可视分析系统，该系统是通过可视化与可视分析的手段，采用图形化的表示方法，提供给用户交互性的操作，让用户可以分析和探索比特币的交易过程。我们可以通过输入块号，然后在交易网络中显示出该块所包含的详细的交易记录。

4.1 地址合并策略

出现在当前数据中的每个地址对应一个节点（黄色）。每个可能的用户也被标示为一个节点，且用颜色区分于地址节点（绿色）。节点的大小和节点的总收入正相关。因为绝大多数情况下每次交易如果有多个输入地址，输入地址往往属于同一个钱包（即属于同一个人），所以我们认为可以把满足这样的条件的节点合并。合并的方法是将这样的一组地址节点作为同一个用户节点的子节点。当两笔或以上的交易分别对应的输入包含至少一个相同的地址时，将这多组输入合并，表现为同一个用户的行为。选择区块之后，默认显示的是用户节点，通过点击用户节点可以做到展开显示对应的各个地址节点。

4.2 合并地址的意义

合并地址之后图的表达将会更加简单。因为经常有一个交易包含多组输入的情况，将地址合并之后，默认状态下图中需要显示的节点数目被有效地减少了。在默认状态下（即不展开的前提下），一笔交易如果输入地址n1个，输出地址n2个，则最多向图中添加(1 + n2)个新节点。而如果不进行合并，则最多需要添加(n1+n2)个新节点。从数据特征观察得出的一个结论是，涉及大额比特币的交易因为单个地址存储的比特币没有达到输出要求等原因，经常涉及多个输入地址。我们感兴趣的交易经常含有很大数量的输入地址。进行合并之后，图的表达变得高效许多。

合并地址使得图的表达变得合理。如果完全不合并地址，即，每一笔交易之中的每一个地址都对应一个新的节点，则图中将出现数目相当庞大的冗余，且交易的输入输出的关系将难以合理表述，需要引入虚拟节点来合理表述这样的输入输出流。如果我们合并地址相同的节点，使得每个地址对应于唯一一个节点，上述表述混乱的情况甚至会加剧。然而，如果我们明确每个输入地址和输出地址都对应到一个用户，然后用图表示用户之间的输入输出关系，则图的表示会变得清晰合理。我们可以很容易看出每个节点和每个边的含义，这也有利于后续的交互分析。

合并地址有利于分析。当我们不合并地址时，地址之间的关联关系很难被归纳出来。当我们将地址之间"很可能归属于同一个人或者同一个组织"的关联表达出来之后，一些交易行为之间的关联就变得更加醒目，一些之前没有充分表达出来的关联也能够被表达出来。

我们合并地址的方法可能有过度的合并，即，将不属于同一个人的两个地址合并为"可能属于同一个人"的地址。这是因为，在极个别的案例里，有可能有两个或者以上不同的人因为某些原因互相信任，共同参与同一笔交易的支付。在以往的研究中，这样的案例被认为是可以忽略不计的极少数，我们通过与比特币矿工的交流也认为这样的情况极少发生。事实上，主要的问题其实是我们的合并尚且不足以将属于同一个人地址归纳完全。合并不足才是最主要的问题。

4.3 交易网络

每个节点表示一个合并后的实体，每个实体包含至少一个地址。节点的大小表示该节点所合并的地址的数目。交易之间的通过地址联系起来，我们可以看到交易之间的关系和交易的传递关系，帮组我们来发现一些有趣的现象。然后用户可以通过拖拽操作看到一些感兴趣的交易。通过点击每个节点可以展开或者合并每个实体的地址，能够让我们看到更加丰富的信息。

4.4 信息提示

在屏幕的右边显示信息提示栏。展示了块信息和交易节点信息。块信息包括了块的高度，所包含的交易费，所包含的交易的数目和交易的时间。交易节点信息包含了，节点编号，输入，输出费用和交易时间。我们可以交互性的在交易网络视图中选择感兴趣的交易节点，然后能够看到详细的有节点信息。

5 案例分析

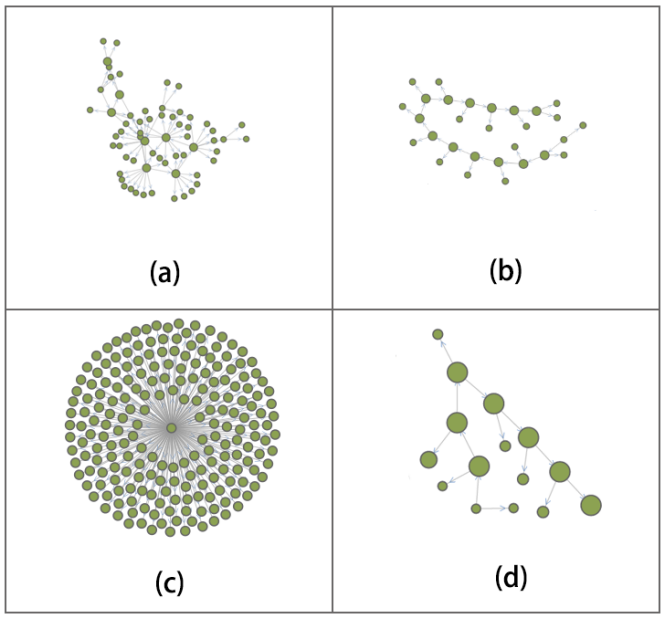

Fig. 2. 案例分析(a)交易社区，(b)传递交易单链，(c)一对多的交易，(d)大额交易的传递

5.1 交易社区

在短时间之内形成彼此之间相互关联的复杂交易关系，也是值得注意的情况。这种情况几乎不可能发生在交易双方彼此陌生、偶然发生交易的情况。同样以高度200000的区块为例，包含地址13Pc1mxKmSWnFiHhzcFo2SWwVsFum2KyYs的子图显示出了高度的复杂性。虽然每个节点涉及的输入额度大约仅仅10比特币，但是因为他们在图中形成了相当复杂的关系网络，我们也认为这很可能是一个有组织有目的的活动。

5.2 传递交易单链

众所周知，比特币网络确认交易是需要一定时间的。一般人使用比特币进行交易的时候很少有段时间内连续发生多笔交易的需求。但是在我们对历史数据进行分析的时候发现，例如在2012年8月22日前后，就大规模地出现了这样的特殊交易特征。以高度200000的区块中记录的交易信息为例，包含地址1313uJwxieFUdZ1p8uhdtcmmCPBtdXxagm的长链虽然每笔交易数额并不庞大，但是彼此之间相互勾连，形成了复杂的长链。考虑到每个区块中记录的仅仅是大约10分钟左右的数据，这些地址的拥有人即使不是同一个人，也有极大几率属于一个成员之间彼此高度信赖的组织。而他们进行这样的交易的目的也是值得我们进一步探究的。

5.3 一对多的交易

仍然以高度200000的区块为例，包含地址13h1DP2Boo9TAsenphroACxhNy7pGxDYXd的子图形成了一个很有意思的中心包围图形。我们认为拥有地址13h1DP2Boo9TAsenphroACxhNy7pGxDYXd的用户并不是一个普通的用户，因为这个人在短时间之内和许多其他用户发生了交易。这些交易虽然大多额度不大，但是积累起来也为这个值得特别注意的用户带来了267.92948742比特币的入账。如果不合并地址，也许我们就会错过这个特殊的子图。

5.4 大额交易的传递

有些交易虽然不一定形成很长的交易链，但是数额之大已经值得引起关注。在高度200000的区块中，包含了地址16qbYUweJJXtibW4uVftaxSa765XznCzNm的交易链虽然并不是很长（最多只涉及15个用户），但是这个并不算很长的交易链中单个地址在单次交易中的入账就能高达28237.5928635比特币。如此庞大的金额本身就值得重视，更何况还发生了成链的现象。这些各不相同的奇特子图存在于同一个区块中，我们认为可能不是偶然的。有可能是在那段时间里，某个组织在进行洗钱活动。这样的洗钱活动经过再深一步的分析，也许可以追溯到早先的某一笔或者某几笔大额交易。

6 讨论及总结

虽然合并地址能够有效地提高表达效率，使得图逻辑清晰，并且有助于发现交易模式特征，但是这样的合并结果使得所有的节点看起来只有两类：地址节点和用户节点。任何两个地址节点和用户节点之间除了大小不同以及关联关系之外看起来都是相同的，除非通过单击对应节点查看详细内容。事实上，比特币历史数据记录的基本单位是交易，而这种图的表达强化了从交易中总结出的“关系”信息，却弱化了交易本身的一些其他信息。比如，交易发生的先后顺序很难充分表达出来。甚至一个地址是输入还是输出多一些，也很难从图中得到直观印象。用点的大小表达收入总额，则很难从图中直观地看出哪个用户拥有很多地址；用点的大小表达每个地址参与的交易数或者每个用户拥有的地址数，则很难直观表达交易的额度。在实际应用中，人们既关注交易额度，也关心交易数目。参与巨额交易的用户值得关注，一个参与大量少额交易的用户同样是值得重点关注的。两个因素相比之下，一般认为交易额度的因素更加重要。这就是为什么本文选择交易额度来对应于节点的大小。我们将采用把多个因素耦合成可以直观表达的变量，如大小、颜色深浅等，以便达到较为直观地表达的目的。对于所显示的交易我们应该根据时间分隔而不是根据块来分割。当所需显示交易数目较多的时候，整个图的结构会过于庞大，我们应该只显示我们感兴趣的重要交易。

目前，我们已经获取到了四分之三的历史交易数据。接下来我们可以对数据作进一步的分析。可以从一个大额交易出追溯比特币的流向，我们还可以进一步分析一些著名的交易事件，试图去发现一些异常的交易信息。在分析比特币的交易的时候发现不少的交易被拒绝，我们应该特别关注这些交易产生的原因，并给出适当的解释，帮组我们进一步的理解数据，揭露交易背后隐藏的信息。

REFERENCES

[1] D. Battista, P. Eades, I. G. Tollis, and R. Tamassia. *Graph drawing: algorithms for the visualization of graphs*. 1999.

[2] G. D. Battista, V. D. Donato, M. Patrignani, M. Pizzonia, V. Roselli, and R. Tamassia. Bitconeview: visualization of flows in the bitcoin transaction graph. In *Visualization for Cyber Security (VizSec), 2015 IEEE Symposium on*, pp. 1–8, Oct 2015.

[3] R. A. Becker, S. G. Eick, and A. R. Wilks. Visualizing network data. *Visualization and Computer Graphics, IEEE Transactions on*, 1(1):16–28, 1995.

[4] U. Brandes, G. Shubina, and R. Tamassia. *Improving angular resolution in visualizations of geographic networks*. Springer, 2000.

[5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*, vol. 6. MIT press Cambridge, 2001.

[6] P. Eades. A heuristics for graph drawing. *Congressus numerantium*, 42:146–160, 1984.

[7] M. Fleder, M. S. Kester, and S. Pillai. Bitcoin transaction graph analysis. *arXiv preprint arXiv:1502.01657*, 2015.

[8] A. Frick, A. Ludwig, and H. Mehldau. A fast adaptive layout algorithm for undirected graphs. In *Graph Drawing*, pp. 388–403. Springer, 1994.

[9] T. M. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

[10] I. Herman, G. Melançon, and M. S. Marshall. Graph visualization and navigation in information visualization: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 6(1):24–43, 2000.

[11] D. Holten and J. J. Van Wijk. Force-directed edge bundling for graph visualization. In *Computer Graphics Forum*, vol. 28, pp. 983–990. Wiley Online Library, 2009.

[12] T. Kamada and S. Kawai. An algorithm for drawing general undirected graphs. *Information processing letters*, 31(1):7–15, 1989.

[13] A. Lopez, B. Alvarez-Pereira, S. Gorsky, and M. Ayres. Network and conversation analyses of bitcoin. 2014.

[14] M. S. Marshall, I. Herman, and G. Melancon. An object-oriented design for graph visualization. *Software: Practice and Experience*, 31(8):739–756, 2001.

[15] T. Munzner. Drawing large graphs with h3viewer and site manager. In *Graph Drawing*, pp. 384–393. Springer, 1998.
[16] T. Munzner, E. Hoffman, K. Claffy, and B. Fenner. Visualizing the global topology of the mbone. In *Information Visualization'96, Proceedings IEEE Symposium on*, pp. 85–92. IEEE, 1996.
[17] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008.
[18] F. Reid and M. Harrigan. *An analysis of anonymity in the bitcoin system*. Springer, 2013.
[19] D. Ron and A. Shamir. Quantitative analysis of the full bitcoin transaction graph. In *Financial Cryptography and Data Security*, pp. 6–24. Springer, 2013.
[20] G. J. Wills. Nicheworks—interactive visualization of very large graphs. In *Graph Drawing*, pp. 403–414. Springer, 1997.